1

Concepts and Constraints of Habitat-Model Testing

MELVIN L. SCHAMBERGER and L. JEAN O'NEIL

Abstract.—Habitat models used in land-use-planning studies are based on assumed relationships between the animal and its environment, and between habitat suitability or capability and some measure of the animal. Attempts to validate habitat models have had mixed results. Some of the problems are related to the delineation of the components of habitat as contained in the model, and others are related to the difficulty of obtaining an independent measure that can serve as a suitable standard of comparison. This chapter briefly presents the conceptual basis for Habitat Suitability Index (HSI) models being developed by the U.S. Fish and Wildlife Service, discusses examples of model validations, and proposes methods for testing models as they relate to land-use applications.

The use of quantitative habitat models for fish and wildlife planning is relatively new, but is increasing rapidly for inventory, impact-assessment, mitigation, and wildlife-management studies (Thomas 1982; Urich and Graham 1983). Models provide a needed tool for studies where habitat is to be emphasized in the natural resources planning and decision-making process. Many types of models are collectively referred to as "habitat models," and most of these are used in an attempt to record or predict a species' response to its environment; this response may be described as occurrence, physiological condition, abundance, or other response of interest to the model user. The response of interest then becomes the model objective, and differing objectives must be recognized in model construction and testing.

Habitat has been selected as the basis for modeling efforts to assist in planning studies because habitat provides an integration between concepts of population and carrying capacity, and it can provide a consistent basis for baseline, impact-assessment, mitigation, and monitoring studies (Fish and Wildlife Service 1980a, 1980b). Habitat Suitability Index (HSI) models are being developed and used in the context of determining habitat quality (Fish and Wildlife Service 1981a). Such models usually are consistent with data needs for planning studies; they are fairly simple, can be applied in a timely manner with minimum cost, and the outputs are easily understood. The reliability of these planning models is not as high as we would like, yet alternatives have not been developed that are widely accepted. Science and planning are two separate realms (Romesburg 1981). Science requires high confidence levels, precision, and certainty. The standards of science are not applied to planning decisions that merely require that decisions be made, using the best available information. How can we integrate the concepts and rigor of science into the realm of planning? Perhaps habitat models can provide a link.

HSI model validation is difficult for several reasons. There are no standard methods for defining or measuring habitat quality. We often lack quantitative habitat data from which to develop models and data that exist are not available in a consistent format. Further, the models are developed around the concepts of habitat and carrying capacity, terms that have no commonly accepted definitions and that are difficult to quantify. Conversely, models are often tested around the concepts of population dynamics and habitat use, and testers have not always recognized the conceptual differences between habitat and population models.

This chapter provides a brief discussion of the conceptual relationships among habitat, carrying capacity, and habitat models as they relate to model testing. The focus is primarily on HSI models, although the concepts, constraints, and suggestions for testing are generally applicable to most types of habitat models. Several approaches to model testing are presented, with the overriding caution that models must be tested and applied within the context of model objectives and their intended use.

Habitat model concepts

Models are simplifications of the systems they depict, and they lose resolution in attaining simplicity. They require numerous assumptions and can never completely mimic the real world (Maynard Smith 1974; Hall and Day 1977). If the simplification is done properly, most system dynamics are preserved (Maynard Smith 1974) and extraneous information is removed from further consideration (Overton 1977). Models provide a framework around which qualitative habitat information can be structured for decision making and can structure qualitative and quantitative relationships into testable hypotheses.

Habitat models are based on the concepts of habitat and carrying capacity. Both can be defined either narrowly or broadly (e.g., Edwards and Fowle 1955; Pianka 1974; Giles 1978). Narrow definitions of habitat strictly limit the number

MELVIN L. SCHAMBERGER: Western Energy and Land Use Team, U.S. Fish and Wildlife Service, 2627 Redwing Road, Fort Collins, Colorado 80526

L. JEAN O'NEIL: Waterways Experiment Station, U.S. Army Corps of Engineers, P.O. Box 631, Vicksburg, Mississippi 39180

and type of considerations, such as the basic food, cover, or physical habitat attributes necessary for survival. Broad definitions include other factors, such as competition or abiotic interactions. The principal difference among definitions of carrying capacity is the number of factors perceived to be critical in determining population limits. The most comprehensive view is that carrying capacity is, a function of all factors that interact to limit populations, including food, predators, inter- and intraspecific competition, disease, mortality, natality, weather, and habitat. Thus, habitat is not the only factor that determines animal presence or abundance, and habitat models may include only a few of the factors that determine population levels (Fig. 1.1). It is important to note that each habitat model contains restricted, operational defi-

nitions of habitat and carrying capacity, and those definitions must be considered when designing model tests.

HSI model concepts

HSI models are specifically designed for use in planning and environmental impact assessment studies. They are meant to quantify habitat quality and to permit wildlife resources to be considered along with other aspects of project planning, such as engineering or economics. They are not models of carrying capacity because not all factors that influence animal abundance are included. They are constrained to basic habitat attributes thought to be important both to the wildlife species and to specific planning or management needs. HSI models are designed for use in situations where land use and, therefore, habitat condition are expected to change; they are intended to allow assessments of resultant changes in potential habitat quality and availability for selected wildlife species.

Variables included in the models are limited to those (1) to which the species responds; (2) that can be measured or estimated readily; (3) whose value can be predicted for future conditions; (4) that are vulnerable to change during the course of the project; and (5) that can be influenced by planning and management decisions. Many variables that are known to influence animal populations are excluded from HSI models if they cannot be readily measured (predation), managed (weather), or predicted for future conditions (competition). The result is a model that has a very restricted operational definition of habitat for a specific land-use study and for a specified geographic area.

Many characteristics of vegetation structure meet the above five conditions, e.g., basal area of mast-producing trees or density of the herbaceous layer. Work on multiple species such as the work of Elton and Miller (1954), MacArthur and MacArthur (1961), Hildén (1965), and Wiens (1969) has been important in forming the HSI concept, whereas more recent studies such as Nixon et al. (1980) have contributed to individual species models. In addition to vegetation structure, variables related to floristics (e.g., abundance of preferred foods), interspersion (distance to cover), landform (presence of standing water), and similar factors often meet the above conditions.

The HSI is determined through aggregation of one or more Suitability Index (SI) scores for life-requisite components, such as winter food. By definition, the SI and HSI provide a 0–1.0 index of habitat suitability. The HSI for a species at a given site is not intended to predict population levels, but an HSI of 0.8 should indicate better habitat quality than an HSI of 0.4 and should represent greater *potential* carrying capacity.

In summary, HSI models are not research models, carrying capacity models, population predictors, or comprehensive. HSI models are practical, operational planning models; designed to assess impacts of change; and based on a narrow definition of both habitat and carrying capacity. They also provide a bridge between the fields of planning and science,
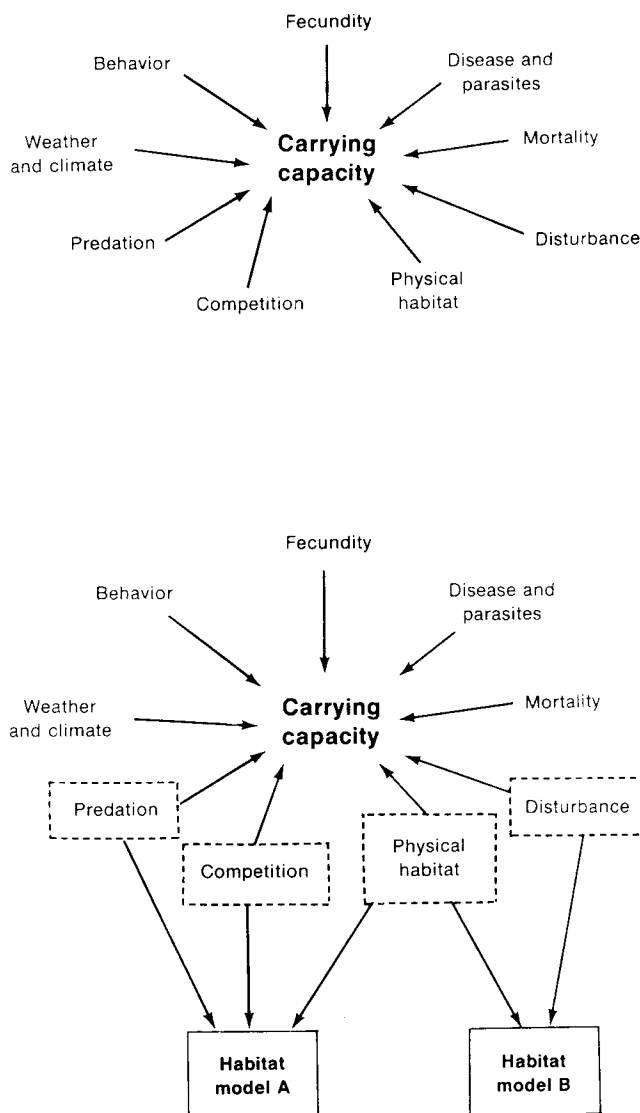

Figure 1.1. Relationships between carrying capacity and habitat models.

in that science is used to improve model performance in planning activities.

Model testing

Model testing serves two important purposes: (1) to provide information about model performance and reliability in specific applications; and (2) to provide data that can lead to model improvement for both the model tested and for similar models. This chapter focuses on the design of tests that will provide meaningful model-performance data. Model tests are intended to determine how well a model meets its stated objectives; therefore, if model objectives are not clearly stated, tests will be nearly impossible to design and results often will be inconclusive. Tests must be developed in consideration of the planned use of the model and should include evaluation of the variables in that model.

Results of model tests must be interpreted in the context of how the model will be applied, because a model may be adequate for one type of study and inadequate for another (Bella 1970). Marcot et al. (1983) supported this concept by referencing several criteria, including precision, accuracy, realism, and resolution. For this reason test results should report the level of confidence or significance attained; a model that performs poorly at the 95% confidence level may be acceptable for some types of operational studies if it performs well at the 75% confidence level.

One of the biggest problems we have encountered in HSI model-testing efforts is that of designing tests and locating either test sites or existing data that are consistent with

model content and purpose. Because each model defines habitat in a slightly different manner, there is no standard approach to model testing. Each of the methods discussed below can be used, under certain circumstances, as the basis for model testing. However, considerable forethought is required to select the appropriate method. Habitat-use data, for example, may be appropriate in one situation, but inappropriate in another. Other considerations in test design are listed in Table 1.1 and discussed below.

LEVELS TO TEST WITHIN A MODEL

One of the first decisions in designing a model test is to determine what part of the model to test. Hypotheses can be formulated at any level (Fig. 1.2) within the model, including tests of assumptions (level A); variables (level B); components (level C); or overall output (level D). Nearly all HSI model tests to date have been at level D, which often provides little information for improving model performance because actual information about the species' response to changing habitat variables is lacking. Overall model performance can be evaluated, however, allowing earlier use of the models in the planning process. Much of the research on wildlife-habitat relationships in the last few years can contribute to tests of assumptions and variables and is a valuable source of information for model construction.

Testing individual model variables (level B) or assumptions (level A) is one approach that provides information for determining and improving model reliability. For example, the species depicted in Figure 1.2 feeds on insects in the tree canopy, but measurement of insect abundance in the tree canopy is difficult. A surrogate measure, foliage density, was substituted for insect abundance in the model, on the assumption that increased amounts of foliage would increase insect habitat, thereby increasing insect abundance. This assumption is a testable hypothesis and was shown by Blenden (1982) and Blenden et al. (Chapter 2) to be correct. Testing at this level also benefits other models that use the same assumption. Level B tests evaluate the relationship between an individual variable and a response by the species (see, e.g., Scott and Oldemeyer 1983), whereas interactions among variables can be tested at level C, in which sensitivity analysis is essential.

The most comprehensive tests should proceed from a testing of assumptions at level A before proceeding to level B, and so forth throughout the model, with the level D test occurring last. Such a thorough test of a model is time-consuming and costly (Caswell 1976b; Overton 1977), but may provide the most productive results in terms of model improvement.

TESTS OF CHANGE

The most logical test is one that evaluates the model within the conditions and purposes for which that model is intended to function. In the case of HSI models, these are planning situations. Because HSI models are designed to help assess habitat changes anticipated from alteration of environmental features, the best HSI model tests are those that evaluate the model or model variables under conditions

Table 1.1. General considerations for HSI model testing

1. Identify model objectives, purposes, and performance levels before designing test. Model tests must be consistent with model objectives and model resolution.

2. Test models against a real-world measure of a response by the animal.

3. Long-term, multiyear data for both habitat variables and animal response should be collected, using stringent quality control of field data.

4. Models should be tested against data sets other than the ones used for model development.

5. Study design should use concepts of change.

6. Tests using exploited species require additional care in data analysis.

7. When variables not included in the model (e.g., predation and competition) are known to be limiting to the test population, correlations between model output and animal numbers probably will not occur.

8. It usually is preferable to test model variables and establish their validity before proceeding to tests of overall model output.

9. Shortcomings of the test data should be fully determined and recognized in reporting test results. Poor correlations between density and model output should not automatically lead to the conclusion that the model is incorrect. The data set may be inaccurate or inappropriate as test data.

10. Large sample sizes are essential; small sample sizes are difficult to interpret and may lead to erroneous conclusions.

11. The standard of comparison should be reflective of habitat quality.

12. Test sites should include the entire range of habitat quality.

13. Test sites should be several times the average home range size of the species being tested.
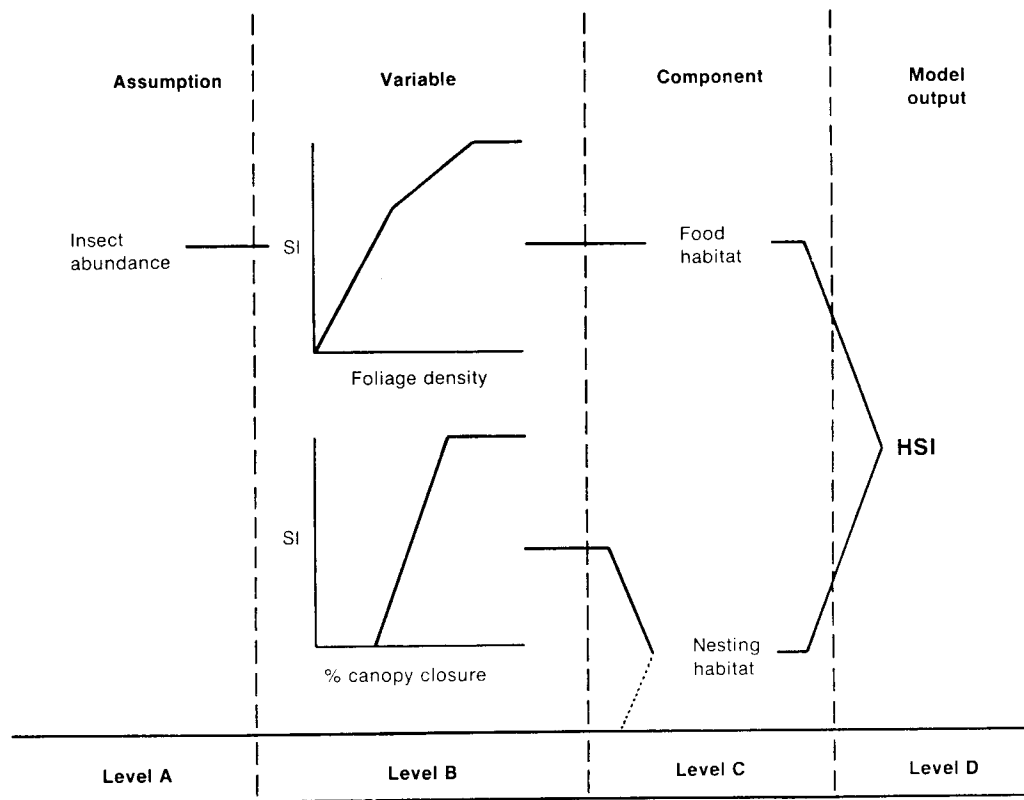
14. Understand the model before testing it.

Figure 1.2. Levels that can be tested within a habitat model.

in which variable values are changed. In such studies, variations in the animal's response (e.g., changes in population levels or reproductive success) can be predicted and directly related to differing values of the variables.

Tests employing change can use a variety of test methods; the experimental design of the test is the critical element. There are three general categories of tests of change. The first includes laboratory or field tests where, under controlled conditions, different variable values are related to animal response (e.g., Wiens and Rotenberry 1981b). For example, grass composition and height might be manipulated and the response by nesting waterfowl determined.

The second category encompasses situations in which one uses different existing vegetation conditions to represent change in the values of interest. For example, O'Neil is presently testing a pine warbler (*Dendroica pinus*) model on sites that display a range of variability in tree density, one of the model variables. Current work on vegetation-succession models (e.g., Smith et al. 1981a; Raedeke and Lehmkuhl, Chapter 53) is expected to improve the sophistication and usefulness of this type of test. An alternative approach is to select sites with a range of values for the animal response (e.g., sign, density) and then compare observed scores with scores predicted by the model. A drawback to tests that simulate change is that usually there is insufficient control to determine cause-and-effect relationships and variable interactions, and extreme care must be used in data interpretation.

The third category of tests of changes uses model application and evaluation in operational settings where land-use changes are proposed and habitat gains or losses are predicted. The changes are implemented and species' responses over time are monitored and compared to model predictions. We advocate greater use of this type of test.

STANDARD OF COMPARISON

Once the structure of a test has been determined, an appropriate standard, such as habitat use, density, or condition factors, must be selected for comparing the output of the model or model component. The standard of comparison is selected after consideration of characteristics of the species; model variables, objectives, and assumptions; field conditions and season; data used in model construction; and the reliability of the potential standard.

Theoretically, a model should be developed and tested using two or more different sets of the same type of data. Realistically, most terrestrial HSI models are developed from several sources using different types of data, such as habitat preference, reproductive success, or lethal condition (e.g., winter temperature). The model, by use of suitability-index graphs (Fig. 1.3), converts these data to a common 0–1.0 index format. However, when model output is tested, often only one standard is used for comparison; inadequate consideration is given to matching the test data with the data used for curve and model development. For example, if lethal-condition data were used to develop a model, perhaps
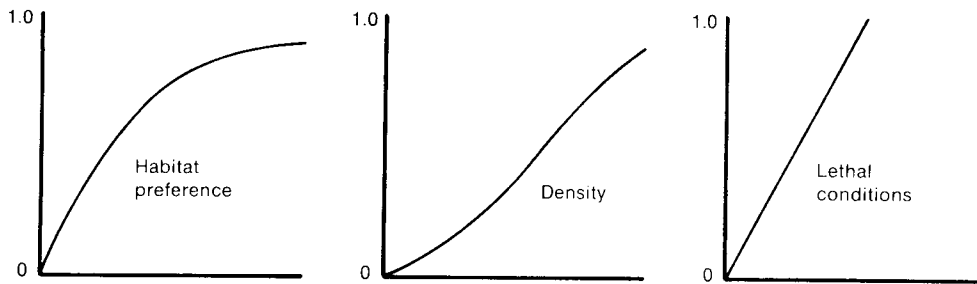
Figure 1.3. Types of data commonly used in models that must be converted to a 0–1.0 scale for model output.

presence/absence data could be used as a test statistic, whereas continuous-measurement data such as animal density might not be appropriate. Conversely, a model designed to assess potential presence/absence conditions should not be expected to predict population densities over a wide range of habitat values.

There is no simple solution to this problem, although we can urge our colleagues to gather and report species' habitat data in a common and consistent format that can be translated into suitability curves for habitat models. In the interim, it is incumbent upon model builders to identify clearly the type(s) of data used as the basis for curve development. In this way, model testers will know the assumptions used in the model and can design tests accordingly. Until consistent data are available, we may need several models for the same species, each based on different types of habitat data and designed for a particular use. The tests must recognize these differences in models.

Tests addressed in this chapter are those that relate model output to measures of the "real world" and can be categorized into (1) habitat-use information; (2) abundance data, such as standing crop or population indices; or (3) measures of well-being, such as growth, reproductive success, or endocrine levels. Regardless of the measure used, the measure itself must reflect habitat quality; otherwise that standard of comparison is inappropriate. Model tests are complicated because the independent measures used to define habitat quality have unique sets of assumptions that create problems in data use, reliability, and interpretation (see Miller 1984).

Habitat use

Habitat-use data document the species' use of or preference for particular areas within its range. It is assumed that (1) a species will select and use areas that are best able to satisfy its life requirements; and (2) as a result, greater use will occur in higher-quality habitat. Numerous techniques are available to assess habitat use, including direct observation, animal tracking, call count, and radio telemetry. For example, Lancia et al. (1982) used radio-tagged bobcats in the Southeast to determine whether habitat use by tagged bobcats correlated with HSI model predictions. The two assumptions mentioned above are not always valid, because factors other than habitat characteristics may affect an animal's use of a site or our perception of that use. The researcher must demonstrate that a measure of use and habitat quality are related. Further, species that have low population levels unrelated to habitat limitations are difficult to test with these techniques (see Part III of this volume). Species with low mobility, specific habitat requirements, and high detectability are more suited to this type of habitat-quality measure. Laymon and Barrett (Chapter 14) discuss some particular problems with this approach.

Animal-abundance data

Animal-abundance data include complete animal counts, indirect counts, and various indices. Three factors may confound the use of abundance data as a standard of comparison: (1) density may not be a function of habitat quality in the study area; (2) density is difficult to measure accurately; and (3) not all factors limiting the population are incorporated in HSI models.

Population levels do not necessarily reflect habitat quality, as Van Horne (1983) showed. Population determinants such as those in Figure 1.1 may override habitat features, or variation in animal numbers may be explained by considering the scale of measurement or stochastic factors (Rotenberry, Chapter 31). Van Horne (1983) provided examples in which density may be higher in low-quality habitat and vice versa because of social interactions. Population levels of many species are often determined at locations or times of the year other than those that are the subject of a model (e.g., Fretwell 1970). Further, point-in-time or short-term population studies reflect only the recent past and may reflect long-term abundance inadequately. Van Horne (1983) recommended consideration of several factors in combination, including density, immigration, and reproductive success, to better link relationships between habitat quality and populations.

The second confounding factor is that the reliability of all population data is often low or uncertain. For example, some individuals or species have varying responses to capture or observation attempts, such as "trap-happy" small mammals or wary small birds. Harvest data are subject to vagaries such as a change in hunting effort, weather, or market prices. Some species experience cyclic changes in population densities, both over seasons and over years, and such cycles are not always habitat-related. In addition, established techniques for gathering population data may be unreliable or may be applied in an unreliable manner. Sources of

error include observer ability and consistency, weather conditions, animal detectability, gear efficiency, and other factors (Miller 1984).

The third confounding factor relates to the fact that HSI models do not incorporate all birth- and death-rate factors that influence abundance. Habitat models usually do not include the direct impact that competitors, predators, parasites, prey, or exploitation may have on populations, even though these factors may strongly influence animal abundance. For example, exploitation is not usually included as a habitat variable in HSI models, yet the results of exploitation may reduce population levels, alter standing crop estimates, change behavior patterns, or force animals into or out of specific habitats.

HSI tests using population data include Clawson et al. (1984) in Missouri, Cole and Smith (1983) in West Virginia, and Clark and Lewis (1983) in Georgia. In the latter studies, the authors used data from a single year and found it inadequate, a point well illustrated by Rice et al. (1983b). Clark and Lewis also found a mismatch between variables in the model and the limiting factors; hence, the model did not predict animal density. Relative abundance is best used as test data when (1) a large number of sample sites are included; (2) the data represent long-term abundance; (3) unreliable data can be screened, as, for example, when unusual weather locally affects one site but not others; (4) field methods are consistent over sites, techniques, observers, seasons, and times of day; and (5) the model and test data are for the same geographic area. This discussion is not intended to conclude that density or animal-use data should not be used as model-test data. Rather, it is intended to point out that density or use data alone will not always provide a valid standard of comparison for HSI model testing. Alternative measures could include reproductive success or condition factors.

MEASURES OF WELL-BEING

Measures of well-being or condition factors reflect the state or health of an individual or population and are assumed to reflect habitat quality. Van Horne (1983) felt that, in some instances, factors such as reproductive success or mean body weight may be better indicators than density of habitat quality. These measures pertain to vigor, growth, reproductive success, stored energy, or endocrine system products that affect metabolism or other features such as blood chemistry, blood cell counts, serum proteins, and enzyme levels. Measurements of well-being have not been used for HSI model tests to date, but they have been used in wildlife management (Kie et al. 1983); Norman and Kirkpatrick 1981; Weber et al. 1984) and are discussed here to stimulate thought regarding their use as standards of comparison. Measures of well-being or condition factors should be used in conjunction with population data. Additional research is needed to identify parameters that correlate with habitat suitability (e.g., Grue et al. 1983; Weber et al. 1984).

Acknowledgments